

# PRONUNCIATION BASED TRANSATER

Final project

딥러닝 2조



# Tabel

- **INSIGHT**
- **TOPIC**
- **PRIOR STUDY**
- **MODEL AND DATASET**
- **FINAL MODEL**
- **USE-CASE EXAMPLE**

# INSIGHT



- While passing by, the phrase “깜 언” was overheard.
- Its meaning was not immediately understood.
- However, the utterance had already concluded before clarification could be sought.
- Nevertheless, the semantic intent behind the expression remains a subject of curiosity.

# INSIGHT

## INPUT

영어 : 땅큐

일본어 : 아리가또 고자이마스

베트남어 : 깜 언

중국어 : 셰 췬

이탈리아어 : 그라찌에

프랑스어 : 메르시

스페인어 : 그라시아스



## OUTPUT

Pronunciation  
based  
Translator

감사합니다

# INSIGHT : WHY JAPANESE???



- In the case of English, typing is relatively convenient
- Japanese is difficult to type on digital devices



- Japanese includes a complex writing system based on Chinese characters, it is fundamentally composed of phonetic scripts such as Hiragana and Katakana. Like English and Korean, it has a structure where words are written based on how they sound.

# TOPIC

(나닌 데스? 무슨말이지)

난닌 데스카?



# TOPIC

## INPUT

아리가토오 고자이마스

아리가또 고자이마스

아리가토 고자이마스

아리갓또 고자이마스

아리갓또오 고자이마스까



## OUTPUT

Pronunciation  
based  
Translator

감사합니다

# PRIOR STUDY

01

## Machine Transliteration (Knight & Graehl, 1998)

Pronunciation-based character mapping is required for conversions such as Romaji → Hiragana, Romaji → Hangul, or Hangul ↔ Hiragana.

Angela Johnson	New York Times	ice cream	
アンジラ・ジョンソン	ニューヨーク・タイムズ	アイスクリーム	
(a n jira jyo n so n)	(nyu u yo o ku ta i mu zu)	(a isukurii mu)	
Omaha Beach	pro soccer	Tonya Harding	
オマハビーチ	プロサッカー	トニヤ・ハーディング	
(omahabiitchi)	(purosakkaa)	(toonya haadingu)	
ramp	lamp	casual fashion	team leader
ランプ	ランプ	カジュアルヒアッション	チームリーダー
(ranpu)	(ranpu)	(kajyuaruhashshyon)	(chiimuriidaa)

Try Pitch

02

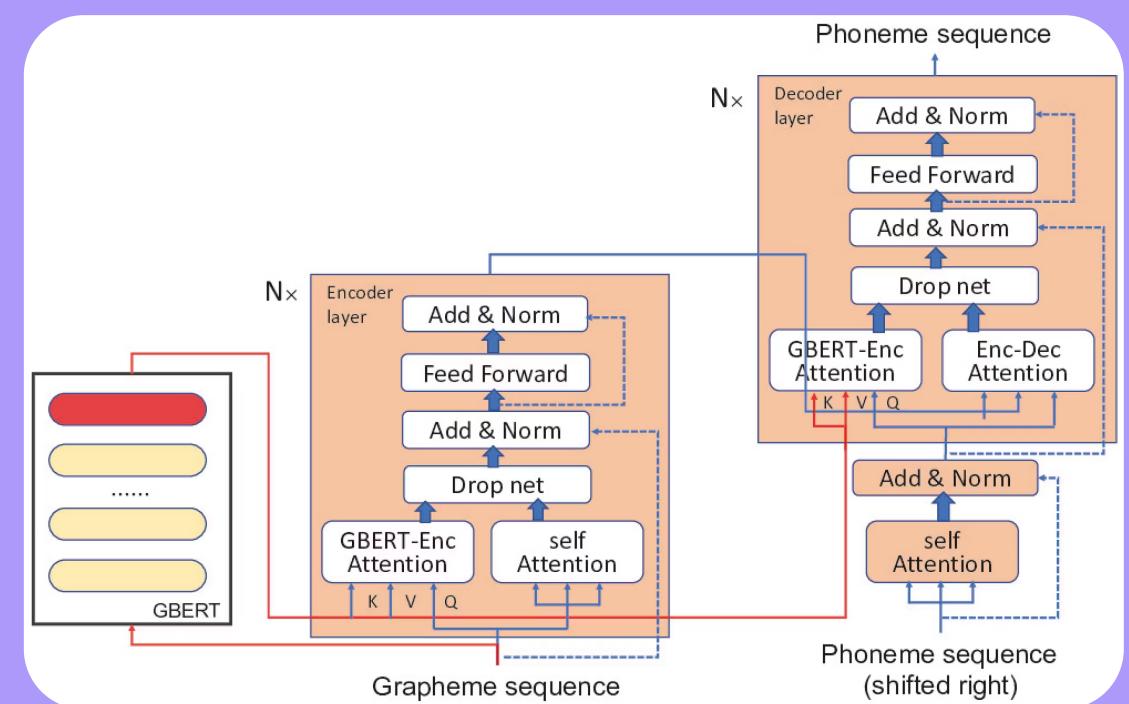
## Umoren et al. (2015) "Grapheme-To-Phoneme, Phoneme-To-Grapheme (G2P-P2G) Transcription Machine for Tone Languages: A Database Approach"

Unlike standardized phonetic symbols (such as SAMPA), which do not allow arbitrary input, this project adopts an arbitrary Korean pronunciation input method.

03

## A Survey of Grapheme-to-Phoneme Conversion Methods (Appl. Sci. 2024)

(Pronunciation → Converting to Characters).



# **MODEL**

# **AND**

# **DATASET**

# Original Dataset

- **Collect Japanese sentences**  
→ Collect sentences from internet  
(240000 sentences)

1297,jpn,きみにちょっとしたものをもってきたよ。  
4702,jpn,何かしてみましょう。  
4703,jpn,私は眠らなければなりません。  
4704,jpn,何してるの？  
4705,jpn,今日は6月18日で、ムーリエルの誕生日です！  
4706,jpn,お誕生日おめでとうムーリエル！  
4707,jpn,ムーリエルは20歳になりました。  
4708,jpn,パスワードは「Muiriel」です。  
4709,jpn,すぐに戻ります。  
4710,jpn,知らない。  
4711,jpn,何と言ったら良いか分かりません。  
4712,jpn,きりがない。  
4713,jpn,何と言ったらいいか・・・。

# Original System Architecture

- 1** **Korean Pronunciation → Romanized text**  
아리가토 고자이마스 → aligato gojaimaseu

**korean-romanizer**

- 2** **Romanized text → Japanese text**  
aligato gojaimaseu → ありがとうございます

**seq2seq**

- 3** **Japanese text → Korean meaning**  
ありがとうございます → 감사합니다

**mbart50**

# Problem

- Seq2Seq's bleu score: 0.5  
noise is too strong to fit  
mbart model
- M-bart fine\_tunning →  
but result was not good

발음을 입력하면 의미를 예측합니다.  
발음을 입력하세요: 아리가토 고자이마스  
발음법칙 적용: 아리가토 고자이마스  
로마자 변환: aligato gojaimaseu  
일본어 복원: シテガと じじ ゃ い ます。  
한국어 번역 결과: 시에요.

# Result



Corpus BLEU: 0.05

# New System Architecture

## 1: [일본어 한글발음 - 해석] DATA SET

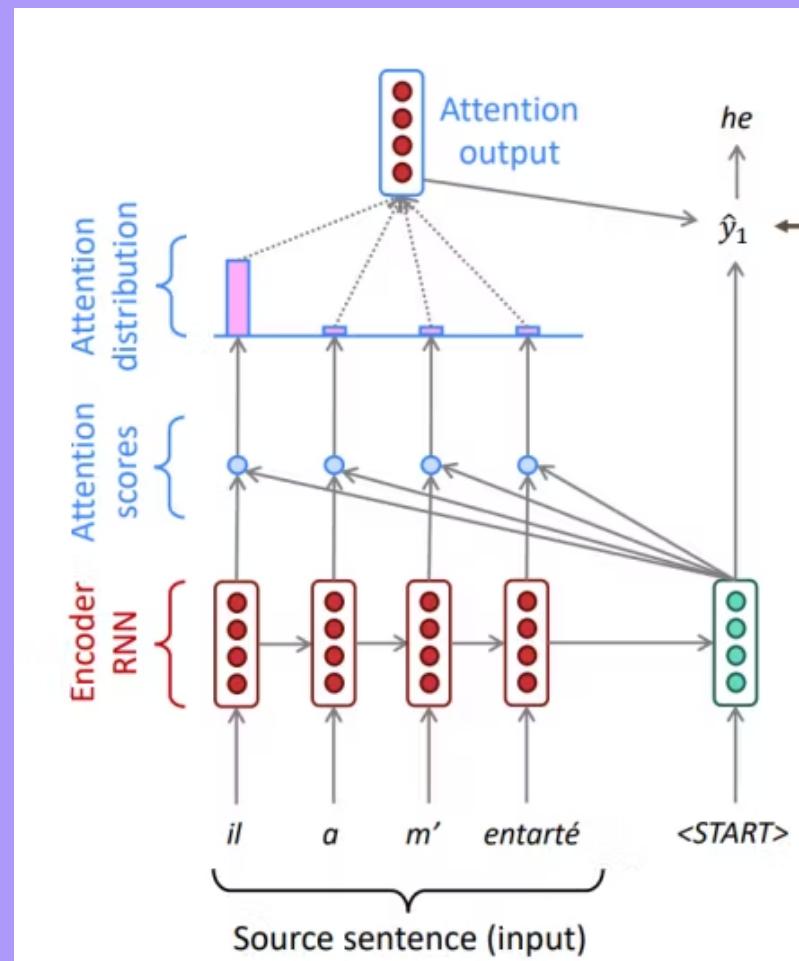
"아리가토 고자이마스"  "감사합니다"

## 2: SEQ\_TO\_SEQ MODEL LEARNING

Build a Simplified Translation MODEL

# New System Architecture

## Seq2Seq with attention



- followed the core architectural design and parameter settings suggested in the original paper

- applied some lightweight modifications  
( reduced embedding and hidden dimensions)

- used bleu score for evaluation

# New Dataset

1

## Collect Japanese sentences

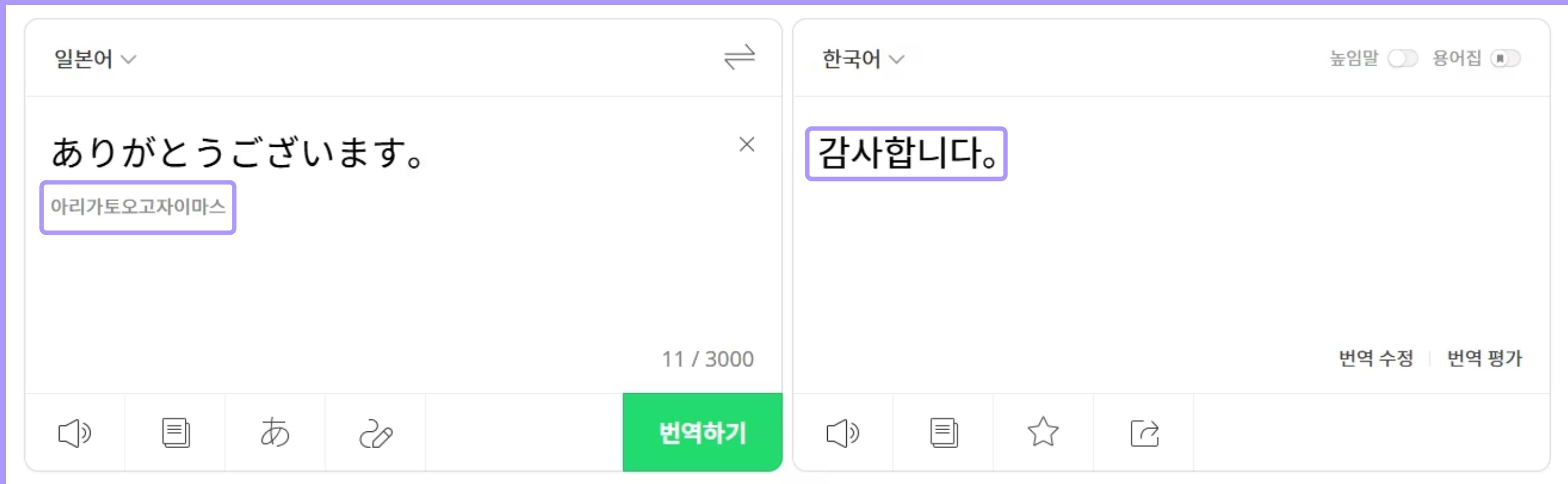
→ Collect sentences from internet (240000 sentences)

2

## Crawl Korean proununciation and Meaning

→ Crawled from papago and crawled Korean pronunciation and meaning.

ex)



# New Dataset

1 input, target

2 와타시타치노 혼돈데노 타이자이와 미지카 스기타, 우리의 런던에서의 체류는 너무 짧았다.

3 와타시와 요지니 덴와오 카케나오스 츠모리데스, 저는 4시에 전화를 다시 걸 생각입니다.

4 모오스코시 케메키오 메시야마리마세카, 좀 더 케이크를 드시지 않겠습니까?

5 와타시타치노 혼돈데노 타이자이와 미지카 스기타, 그는 전 재산을 차에 소비했다.

**와타시타치노 혼돈데노 타이자이와 미지카 스기타**

7 코로나카데 시바라쿠 히라카레테이나캇타 미벤토데스코 콘도마는 열리지 않았던 행사입니다?

8 토텁모 요미 푸란데스네오모이데니 노코리소오데스, 마즈 죠오 푸란데스, 츠어으르 나으 거 쟁이오

9 마오지로쿠데 쿠라미 카오오 시타 진부츠닷타, 창백한 **우리의 런던에서의 체류는 너무 짧았다.**

10 벳도오 토토노에루 타메니 다례카 요코시테쿠다사이마스카, 침대를 정돈하기 위해 누군가 보내 주시겠습니까?

11 아나타노 고엔조니 칸샤시테이마스, 당신의 원조에 감사하고 있습니다.

12 코노 짓켄데와 후추우이와 유루사례나이, 이 실험에서는 부주의는 용납되지 않는다.

13 빌오인미키타쿠 나미, 병원 가기 싫어.

14 고가츠니 카레와 추우고쿠니 미키마스, 5월에 그는 중국에 갑니다.

**Pronunciation**

**Meaning**

# Data augmentation

## noising-korean

- **vowel change**

[vowel\_change, prob=1.] 행복한 가정은 모두가 닮았지만, 불행한 가정은 모두 저마다의 이유로 불행하다.

- **phonological\_change**

[phonological\_change, prob=1.] 행복한 가정은 모두가 달랐지만, 불행한 가정은 모두 저마다의 이유로 불행하다.

# Learning process

**Learning models by noise intensity with the preceding augmentation codes (prob = 0.5)**

**The bleu score of the models was good (0.54)**

# Infer result

발음을 입력하세요: 아리가도오 고자이마스

예측 결과: 감사합니다.

발음을 입력하세요: 아리가도 고자이마스

예측 결과: 고고 합니다.

발음을 입력하세요: 아리갓도 고자이마스

예측 결과: 고 있으 있습니다.

발음을 입력하세요: 아리갓또오 고자이마스까

예측 결과: 알일 있을 것입니다.

Original Pronunciation  
→ Correct !

Mis Pronunciation  
→ Incorrect and,  
Cannot Understand

The model struggled to adapt to diverse pronunciation variations and was overly biased toward standardized pronunciation.

# Solution for Overfitting

- Training was conducted by applying varying levels of noise intensity within a single model.
- The dataset (including noise data) was augmented by 2x, 5x, and up to 7x.
- It was determined that using only the previously applied noise method did not resolve overfitting,  
→ Therefore, other types of noise such as random deletion and random substitution were applied.

→ After testing various methods, decided to apply all three approaches (1, 2, and 3) simultaneously.

# Infer result

**Prob = 0.3**

발음을 입력하세요: 아리가토 고자이마스  
예측 결과: 감사합니다.  
발음을 입력하세요: 아리갓또 고자이마스  
예측 결과: 알리 다 있습니다.  
발음을 입력하세요: 아리카또 고자이마스  
예측 결과: 다가 습니다.  
발음을 입력하세요: 아리가또 고쟈이마스  
예측 결과: 감사합니다.  
발음을 입력하세요: 아리다또 고자이맛스  
예측 결과: 다리다 다니다.

**Prob = 0.7**

발음을 입력하면 의미를 예측합니다.  
발음을 입력하세요: 아리가토오 고자이마스  
예측 결과: 감사합니다.  
발음을 입력하세요: 아리가토 고자이마스  
예측 결과: 고고 합니다.  
발음을 입력하세요: 아리가또 고쟈이마스  
예측 결과: 아감사합니다.  
발음을 입력하세요: 아리갓또 고자이마스  
예측 결과: 고 있으 있습니다.  
발음을 입력하세요: 아리갓또오 고자이마스까  
예측 결과: 알일 있을 것입니다.

**Prob = 1.0**

발음을 입력하세요: 아리가토 고자이마스  
예측 결과: 감사합니다.  
발음을 입력하세요: 아리갓또 고자이마스  
예측 결과: 개월요일입니다.  
발음을 입력하세요: 아리카또 고자이마스  
예측 결과: 하나 안 니다.  
발음을 입력하세요: 아리가또 고쟈이마스  
예측 결과: 있습니다.  
발음을 입력하세요: 아리다또 고자이맛스  
예측 결과: 하나와 정말 새가 있어.

# Infer result

랜덤 치환

아리가토 고자이마스 → 아니가토 코자이마스

랜덤 드롭

아리가토 고자이마스 → 아리 () 토 () 자이마스

# Infer result

Prob = 0.3, 0.7 각 1개 | 랜덤 치환 0.2 | 랜덤 드롭 0.2

발음을 입력하면 의미를 예측합니다. (종료: quit)  
발음을 입력하세요: 아리가토오 고자이마스  
예측 결과: 감사합니다.

발음을 입력하세요: 아리가토 고자이마스  
예측 결과: 감사합니다.  
발음을 입력하세요: 아리갓도 고자이마스  
예측 결과: 감사합니다.

발음을 입력하세요: 아리갓또오 고자이마스까  
예측 결과: 감사 .  
발음을 입력하세요:

Original Pronunciation  
→ Correct !

Mis Pronunciation  
→ Correct and,  
Can Understand

→ Incorrect but,  
Can Understand

# Infer result

Prob = 0.3, 0.7 각 2개 | 랜덤 치환 0.2 | 랜덤 드롭 0.2

발음을 입력하면 의미를 예측합니다. (종료: quit)  
발음을 입력하세요: 아리가도오 고자이마스  
예측 결과: 감사합니다.  
발음을 입력하세요: 아리가도 고자이마스  
예측 결과: 감사합니다.  
발음을 입력하세요: 아리갓도 고자이마스  
예측 결과: 감사합니다.|  
발음을 입력하세요: 아리갓또오 고자이마스까  
예측 결과: 감사를 합니다.

**Original Pronunciation**

→ **Correct !**

**Mis Pronunciation**

→ **Correct and,**

**Can Understand**

# Infer result

Prob = 0.3, 0.7 각 2개 | 랜덤 치환 0.2 | 랜덤 드롭 0.2

발음을 입력하세요: 토이레와도코 데스카?

예측 결과: 화장실은 어디 있어요?

발음을 입력하세요: 토이레 도꼬 데스카

예측 결과: 화장실 어디 ?

발음을 입력하세요: 토이레와 도코데스까?

예측 결과: 이는 머어머요.

발음을 입력하세요: 토일레와 도코 데쓰까?

예측 결과: 당희는 어디 있어 안 있

Original Pronunciation

→ Correct !

Mis Pronunciation

→ Correct and,  
Can Understand

→ Incorrect but,  
Cannot Understand

**Too many noises.. This model is not generalizable**

# Final model

# Final Model

## data

Original data

+pronunciation noise(prob:0.3)

+pronunciation noise(prob:0.7)

+random drop noise (아 가토 고 이마스)

+random change noise(아여가토 고가이마스)

**x5 augmentation of original data**

# Final Model

## paramiter

```
BATCH_SIZE = 64
VAL_BATCH_SIZE = 16
EMB_DIM = 128
HID_DIM = 256
EPOCHS = 30
LEARNING_RATE = 0.001
TEACHER_FORCING_RATIO = 0.6
PATIENCE = 5
WEIGHT_DECAY = 0.0001
```

# Final Model- bleu 0.5276

발음을 입력하세요: 토이레와 도쿄 데스카?

예측 결과: 화장실은 어디 있나요?

발음을 입력하세요: 토이레 도쿄 데스카

예측 결과: 화장실 어디에서 살고 있나요?

발음을 입력하세요: 토이레와 도쿄데스까?

예측 결과: 화장실은 어디입니다.

발음을 입력하세요: 토일레와 도쿄 데쓰까

예측 결과: 우리는 어디서 나와 것입니다.

Original Pronunciation

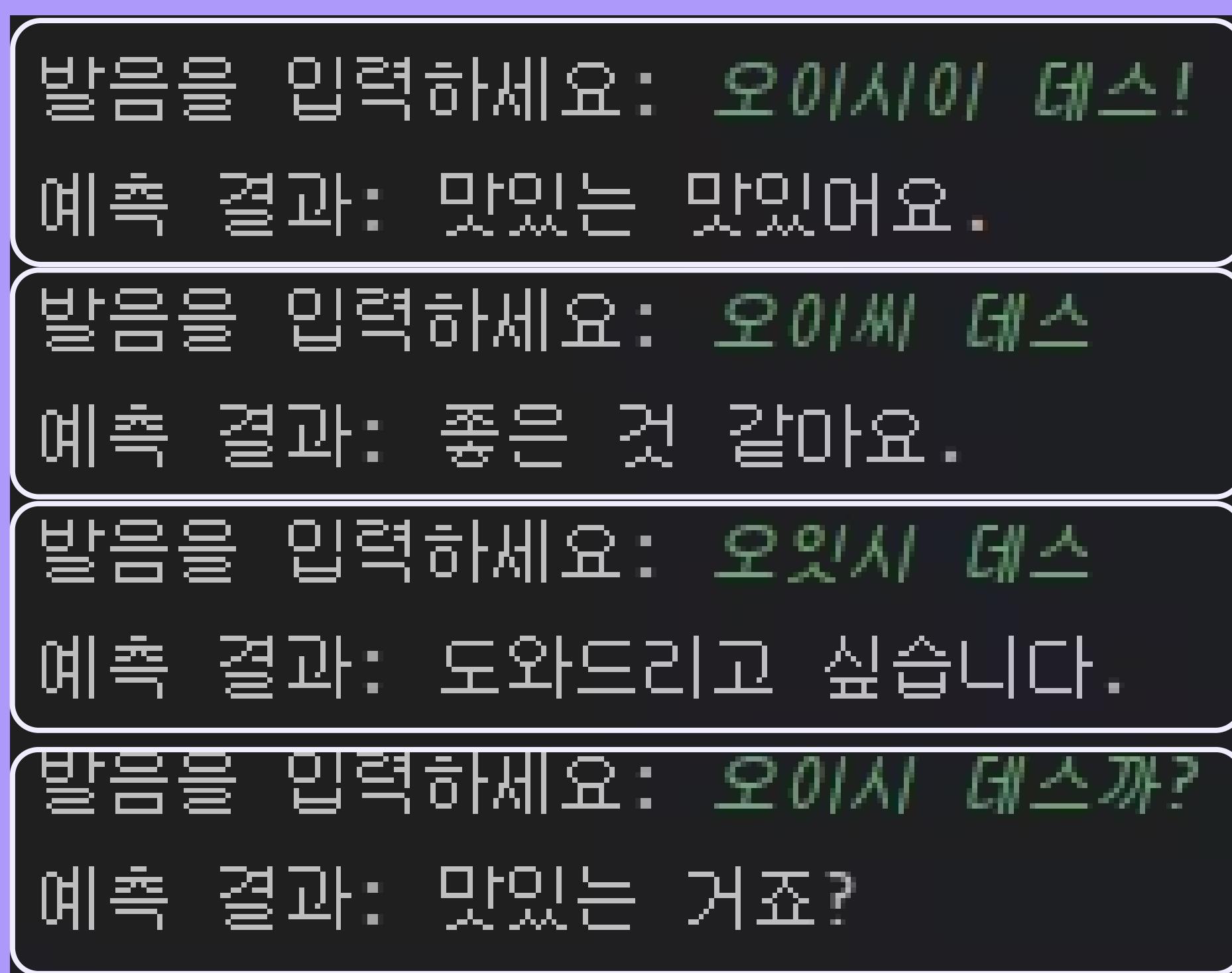
→ **Correct !**

Mis Pronunciation

→ **Incorrect but,  
Can Understand**

→ **Cannot Understand**

# Final Model- bleu 0.5276



Original Pronunciation  
→ **Correct !**

Mis Pronunciation  
→ **Incorrect but,**  
**Can Understand**

→ **Cannot Understand**

# Final Model- bleu 0.5276

발음을 입력하세요: 코노 치카쿠니 콘비니와 아리마스카?

예측 결과: 이 근처에 편의점이 있습니까?

발음을 입력하세요: 코노 치까쿠니 콤비니 아리마스까?

예측 결과: 이 근처에 놀라 있어요.

발음을 입력하세요: 코노 치카쿠니 콘비니 아리마스요?

예측 결과: 이 근처에 편의점이 있어요.

발음을 입력하세요: 고노 치카쿠니 콘비니와 아리마스까?

예측 결과: 5분의 근처에 비비에는 있습니다.

**Original Pronunciation**

→ **Correct !**

**Mis Pronunciation**

→ **Incorrect but,  
Can Understand**

→ **Cannot Understand**

**use  
case  
example**

# Use-case example

(나닌 데스? 무슨말이지)

난닌 데스카?



# Use-case example

(아.. 여기 앉으라는 건가?)

코코니 스와테쿠다사이!



# Use-case example

(우동? 텁푸라? 무슨말이지..  
(물어보기가 좀 그렇네..)

우돈가 데마시타!  
텐푸라오 사키니 메시아캇테 쿠다사이



# Use-case example



발음을 입력하세요: 날님 데스카?  
예측 결과: 몇 명이죠?  
발음을 입력하세요: 코코니 스와테쿠다사이  
예측 결과: 여기 앉아주세요.  
발음을 입력하세요: 우돈가 데마시타  
예측 결과: 우동이 나왔어요.  
발음을 입력하세요: 텐푸라오 사키니 메시아캇테쿠다사이  
예측 결과: 튀김을 먼저 드시오세요.

# Future plans

- 1 train with many paramiter variation
- 2 Enhancing model performance by increasing the amount of training data.
- 3 Compare with other models such as transformer.
- 4 Expanded multilingual capabilities with language detection.

# Pronunciation

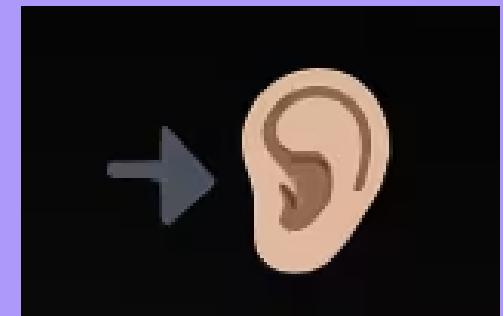
발음을 입력하세요: 이마마데 핫표오오 키이태쿠다삿데 아리가토오고자이마시타

# Meaning

예측 결과: 지금까지 발표를 들어 주셔서 감사합니다. 감사합니다. 감사합니다.

# Q&A

# WHY???



Transcription as *heard*

한국어 ▾

감사합니다

베트남어 ▾

Cảm ơn.

용어집

⇒

×

→

Linguistic speculation

"깜언"도 "깜언"과 마찬가지로 영어 표현 \*\*\*"Come on"\*\*\*을 한국식으로 소리 나는 대로 적은 말입니다. 다만:

- \*\*\*"깜언"\*\*은 영어 발음을 좀 더 가깝게 적은 형태고,
- \*\*\*"깜언"\*\*은 일부 사람들이 장난스럽거나 억양을 강조해서 표현할 때 쓰는 변형된 표기입니다.

둘 다 비슷한 뜻인데, \*\*\*"깜언"\*\*은 더 재밌게, 약간 과장되게 쓰이는 경우가 많습니다.

예를 들어:

- "깜언~ 진짜 이건 해야지!" → "야, 제발 좀 해라~"
- "깜언! 이 정도도 못 하냐?" → "이거 진짜 못 해?" (도발 느낌)

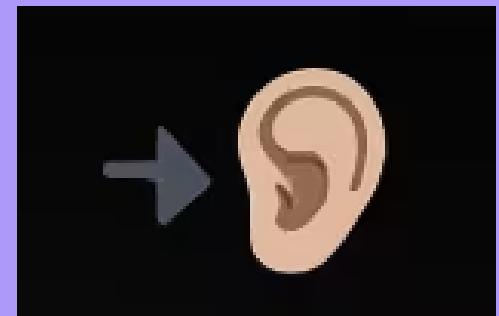
즉, 친근하거나 장난스러운 어투로 재촉하거나 다그치는 말이에요.

요즘 밤처럼 온라인 채팅이나 댓글에서 자주 보일 수 있습니다.

□ ◁ ◁ ◁ ◁ ◁

Foreign Language to Hear << Unexpected Translation Results = Language Trap

# WHY???



Specify language's country

깜언이 뭐야?

"깜언"도 "캄언"과 마찬가지로 영어 표현 \*\*"Come on"\*\*을 한국식으로 소리 나는 대로 적은 말입니다. 다만:

- \*\*"깜언"\*\*은 영어 발음을 좀 더 가깝게 적은 형태고,
- \*\*"깜언"\*\*은 일부 사람들이 장난스럽거나 억양을 강조해서 표현할 때 쓰는 변형된 표기입니다.

들 다 비슷한 뜻인데, \*\*"깜언"\*\*은 더 재밌게, 약간 과장되게 쓰이는 경우가 많습니다.

예를 들어:

- "깜언~ 진짜 이건 해야지!" → "야, 제발 좀 해라~"
- "깜언! 이 정도도 못 하냐?" → "이거 진짜 못 해?" (도발 느낌)

즉, 친근하거나 장난스러운 어투로 재촉하거나 다그치는 말이에요.

요즘 임처럼 온라인 채팅이나 댓글에서 자주 보일 수 있습니다.

□ ▲ □ △ ▣ ▤

아까 깜언이라고 들은거, 베트남 말이였대

맞아요! 아까 말한 \*\*"깜언"\*\*은 베트남어 \*\*"Cảm ơn"\*\*의 한국식 발음이에요.

◆ 정리해서 말하면:

- "Cảm ơn" [깜 언 / 캄 언] → 감사합니다
- 더 정중하게 말하면:
  - "Cảm ơn bạn" → 고마워요 (친구에게)
  - "Cảm ơn các bạn" → 고마워요 여러분 (깜언각반)



# Want to make a presentation like this one?

Start with a fully customizable template, create a beautiful deck in minutes, then easily share it with anyone.

[Create a presentation \(It's free\)](#)